# The diffusion of policy frames: evidence from a structural topic model

Gilardi, Fabrizio ; Shipan, Charles R ; Wüest, Bruno

Abstract: Policy diffusion occurs when policies in one unit (e.g., countries, states, cities) are influenced by the prior adoption of policies in other units. Although numerous studies have convincingly documented this phenomenon, they have, with very few exceptions, generally ignored a crucial step in the diffusion process—namely, how policies are framed ahead of their adoption. Policy frames—the discussion of a policy from particular viewpoints—play a crucial role in linking the actions of previous units with the potential actions in other units. In this paper, we identify policy frames and examine their link with prior policy adoptions. We focus on the area of restrictions on smoking in U.S. states. Our analysis draws upon an original dataset of more than four million paragraphs from articles published in 50 American newspapers covering 47 states between 1996 and 2014. We use structural topic models to estimate how smoking bans have been framed and how frames change as a function of policy adoption in other states. We find that, as more neighboring states enact legislation restricting smoking, concerns about the restaurant business decrease; worries about the casino business increase; detailed regulations such as ventilation requirements or separate rooms for smokers are discussed less frequently; voters' support and involvement in the decision-making process surrounding smoking bans are discussed more frequently; the compatibility of smoking restrictions with individual rights loses salience as a topic; and the passage of legislation is discussed more frequently, while the process by which decisions are made loses salience. These findings establish a foothold for the usefulness of structural topic models and support for the idea that policy frames are an important part of the diffusion process.



Originally published at:
Gilardi, Fabrizio; Shipan, Charles R; Wüest, Bruno (2016). The diffusion of policy frames: evidence from a structural topic model. In: American Political Science Association. Annual Meeting, Philadelphia, 1 September 2016 - 4 September 2016.

# The Diffusion of Policy Frames: Evidence from a Structural Topic Model*

Fabrizio Gilardi[†] Charles R. Shipan[‡] Bruno Wueest[§]

*Work in progress*
August 25, 2016

**Abstract**

Policy diffusion occurs when policies in one unit (e.g., countries, states, cities) are influenced by the prior adoption of policies in other units. Although numerous studies have convincingly documented this phenomenon, they have, with very few exceptions, generally ignored a crucial step in the diffusion process—namely, how policies are framed ahead of their adoption. Policy frames—the discussion of a policy from particular viewpoints—play a crucial role in linking the actions of previous units with the potential actions in other units. In this paper, we identify policy frames and examine their link with prior policy adoptions. We focus on the area of restrictions on smoking in U.S. states. Our analysis draws upon an original dataset of more than four million paragraphs from articles published in 50 American newspapers covering 47 states between 1996 and 2014. We use structural topic models to estimate how smoking bans have been framed and how frames change as a function of policy adoption in other states. We find that, as more neighboring states enact legislation restricting smoking, concerns about the restaurant business decrease; worries about the casino business increase; detailed regulations such as ventilation requirements or separate rooms for smokers are discussed less frequently; voters' support and involvement in the decision-making process surrounding smoking bans are discussed more frequently; the compatibility of smoking restrictions with individual rights loses salience as a topic; and the passage of legislation is discussed more frequently, while the process by which decisions are made loses salience. These findings establish a foothold for the usefulness of structural topic models and support for the idea that policy frames are an important part of the diffusion process.

*We thank Katrin Affolter, Nina Bader, Nina Buddeke, Sarah Däscher, Zander Furnas, Jeremy Gelman, Andrea Häuptli, Geoff Lorenz, Adriano Meyer, Blake Miller, Christian Müller, and Thomas Willi for excellent research assistance, Klaus Rotherhäusler for his help with the implementation of the computational linguistic methods, Fridolin Linder for the evaluation of topic models, and James Cross, Tom Ginsburg, Jude Hays, Katerina Linos, Paasha Mahdavi, Matia Vannoni, Craig Volden, and seminar participants at UC Berkeley, University College London, and the University of Milan for helpful comments. The financial support of the Swiss National Science Foundation (grant nr. 100017_150071/1) is gratefully acknowledged.

†Department of Political Science, University of Zurich (`http://www.fabriziogilardi.org/`).

‡Department of Political Science, University of Michigan (`cshipan@umich.edu`).

§Department of Political Science, University of Zurich (`http://www.bruno-wueest.ch/`).

# 1 Introduction

When political units—states, cities, or even countries—adopt policies, they do not do so in a vacuum, basing their decisions on only internal factors and pressures. Instead, they can observe the actions that other units previously have taken with respect to these policies. Thus, a state that is deciding whether to adopt, say, new gun control laws, or new rules concerning eligibility for various state-funded medical services, can look around to see which other states have adopted such policies, as well as what types of policies these other states have adopted. They can then base their own decisions on what they observe in these other units. This process, known as *policy diffusion*, has been the focus of a large and rapidly growing number of studies (Dobbin, Simmons and Garrett, 2007; Gilardi, 2012; Graham, Shipan and Volden, 2013; Maggetti and Gilardi, 2016). These studies have convincingly established, across a wide range of policy areas, that policies do indeed diffuse, with policies in one unit influenced by policies in other units. That is, these studies have demonstrated that when a unit is considering what to do about a policy, the likelihood that it will adopt the policy is influenced by the existence, in other units, of similar policies.

Although the link between new policy adoptions and earlier policy adoptions has been well established, the focus of the vast majority of studies of policy diffusion has been exclusively on the final adoption decision—that is, did the unit adopt the policy, or did it fail to do so? Although this focus is understandable and has produced numerous important insights, it also ignores a key earlier stage in the policymaking process. In particular, the adoption decision arrives only after policymakers have considered various aspects of the policy. During this stage, the policy can be framed in different ways. Policy frames—the discussion of a policy from particular viewpoints—can shape the final outcomes, including whether to adopt a policy and what form the policy should take (Baumgartner, De Boef and Boydstun, 2008). But policy frames, as part of the diffusion process, can themselves be shaped by the prior policy adoptions that have taken place elsewhere. Thus, a more complete consideration of the interdependence of policymaking needs to account for the link between earlier adoptions and the way policy problems and solutions are defined and understood—that is, how they are framed.

To examine how policy frames change as a function of the adoption of policies elsewhere, we focus on anti-smoking laws—policies restricting or banning smoking in public places—in the United States. Our choice of policy area is motivated by several considerations. First, several American studies (e.g., Shipan and Volden, 2006, 2008, 2014; Pacheco, 2012), as well as abundant anecdotal evidence, indicate that smoking bans have exhibited a diffusion process. This allows us to concentrate on the nature of the process instead of its mere existence. Second, smoking bans have been adopted in a convenient time frame—roughly a ten year period—which is long enough to detect variations and to supply sufficient information but short enough to be practically manageable. Third, the policy has well-defined characteristics and is comparable across units. Fourth, there was significant uncertainty about the potential consequences of the policies along a number of dimensions—economic consequences, popular support, interest group support, ease of implementation, and so on. And finally, this uncertainty over consequences means that the debate over adoption can be framed in multiple ways.

In our empirical analysis we rely on an original dataset of more than four million paragraphs from articles published in 50 American newspapers covering 47 states between 1996 and 2014. More specifically, we use structural topic models (Roberts et al., 2014; Roberts, Stewart and Airoldi, 2016) to identify how these articles have discussed anti-smoking laws and to estimate how these laws have been framed in the states. We then show how these frames change as a function of policy adoption in nearby states. We find that, as more neighboring states enact legislation restricting smoking, concerns about the restaurant business decrease; worries about the casino business increase; detailed regulations such as ventilation requirements or separate rooms for smokers are discussed less frequently; voters' support and involvement in the decision-making process surrounding smoking bans are discussed more frequently; the compatibility of smoking restrictions with individual rights loses salience as a topic; and the passage of legislation is discussed more frequently, while the process by which decisions are made loses salience.

## 2 Policy Frames and the Stages of the Diffusion Process

Policy diffusion occurs if the policy choices of one unit (e.g., countries, states, cities, etc.) are influenced by the policy choices of other units (Dobbin, Simmons and Garrett, 2007; Gilardi, 2012). Although this simple definition captures key elements of the diffusion process, it also omits others. Consider a situation in which State A is deciding whether to adopt a new law. The standard approach, found in most analyses of policy diffusion, is to consider whether State B already has adopted this policy; and then to see whether State B's adoption affects the likelihood that State A adopts the policy.[1] In effect,

[1] Although we refer to "State B," the earlier adoption can be by a single state, as in analyses that examine dyadic relationships between individual states, or by a set of states, as in studies that look at the number of previous adoptions among a specified set of states.

then, these studies implicitly model diffusion as a two stage process; what happens in between these two stages is rarely seen as important.

We argue instead that the process of diffusion occurs in three stages, not two. First, State B adopts a policy. Second, in State A the policy is debated with an emphasis on specific aspects of policy problems and solutions. And third, State A then decides whether to adopt the policy. The middle stage, in which the policy is debated in State A, is more than just a transitional stage; it is worthy of attention in its own right. It is at this stage, when states are considering what to do and debating the policy, that they might consider some of the factors that scholars refer to as the mechanisms of diffusion (Simmons, Dobbin and Garrett, 2006; Braun and Gilardi, 2006; Dobbin, Simmons and Garrett, 2007; Gilardi, 2012). What can they learn about the political or policy consequences of adoptions in earlier states? Would they be likely to suffer negative economic consequences, or would the reap positive economic benefits, if they adopt such a law? Are there norms in place to which they want to adhere, or would they be acting against prevailing norms by adopting a new policy? In other words, especially in this stage, policies can be framed in many different ways. Framing can be defined, quite simply, as "the presentation or discussion of an issue from a particular viewpoint to the exclusion of alternate viewpoints" (Baumgartner, De Boef and Boydstun, 2008, 106).[2]

There are clear links between this second stage, in which a state debates a policy, and the first and third stages. Our focus in this paper is on developing a way to characterize the policy frames that exist at the second stage, and to investigate whether there is a connection between the adoptions in the first stage and the frames in the second stage. But it is worth noting that this connection is important in part because of the link between the second and third stages. This link between the latter two stages is both straightforward and of obvious importance. Put simply, does the way in which an issue is framed within a polity have an effect on the likelihood that the polity will adopt a policy? Especially given that policies usually can be framed in multiple ways, does the specific frame that dominates discussion influence the eventual policy choices? Viewed in this light, policy frames are important as a *cause* of policy outcomes. For instance, Baumgartner, De Boef and Boydstun (2008) have shown that the rise of the "innocence frame" in the American death penalty debate was associated with fewer death sentences.

Our main interest in this paper is instead on the relationship between the first and second stages,

[2]This simple definition is consistent with the one, more detailed, put forward by Entman (1993, 52): "To frame is to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described."

which means that we examine policy frames as an *outcomes*. Given that State A's consideration of an issue is subsequent to State B's action, we investigate whether State B's action influences policy frames in State A. In the area of anti-smoking laws, for example, in one state policies might be framed as being primarily about the health consequences of adoptions restrictions on smoking, while in another debates might concentrate on public support. Does the type of frame change over time? And are frames in a state influenced by the actions taken earlier in other states? In effect, then, our focus is on whether the diffusion process involves policy frames, whereby these frames—which might eventually influence further outcomes—are themselves a product of diffusion from the actions of other actors. Thus, instead of focusing on the direct diffusion from one set of policy outcomes to another, our interest in this paper is in establishing whether previous policy outcomes diffuse to policy frames—a key aspect of the diffusion process that few studies have recognized, let alone examined.[3] In contrast to Baumgartner, De Boef and Boydstun (2008), we do not aim to uncover broad, nation-wide shifts in policy frames. Instead, we are interested in documenting the shifting terms of policy debates at the state level, in a more fine grained way. Moreover, we intend to establish whether, and how, policy frames diffuse as a result of anti-smoking laws becoming more widespread.

To assess which policy frames exist and are most prevalent, and whether the prevalence of these frames is a function of prior adoptions (and thus part of the overall diffusion process), we rely on structural topic models (STMs) (Roberts et al., 2014; Roberts, Stewart and Airoldi, 2016), which we describe in more detail in the following section. This approach allows us to examine, in great detail, which topics dominate the discussion surrounding a policy. Similar to Baumgartner, De Boef and Boydstun (2008), we identify and measure the topics by analyzing media coverage of this policy issue, considering that topics constitute "the smallest units of framing" (Baumgartner, De Boef and Boydstun, 2008, 107). Contrary to these authors, however, we do not examine how the topics might be "used in conjunction with one another to form a larger cohesive frame" (Baumgartner, De Boef and Boydstun, 2008, 136). Instead, our approach focuses squarely on the component parts of frames.

One question that arises is whether the media coverage we examine reflects how policies are framed, or whether it influences the frames. On this question we are agnostic. Regardless of whether this coverage reflects or influences frames, media coverage can be used as an accurate source for identifying the ways in which smoking bans are framed and, more generally, "as an indicator of the nature of public

[3]A notable exception is Pacheco's (2012) study, which not only examines how prior adoptions influence public opinion, but also investigates whether these changes in public opinion then influence adoptions.

discussion" (Baumgartner, De Boef and Boydstun, 2008, 20). Consequently, we use the information derived from structural topic models both to identify the most common frames and to identify their distributions, both cross-sectionally and over time. In the analysis in this paper, we will look specifically at whether the prevalence of specific topics is a function of adoptions in other states—that is, whether policy frames change as a function of the adoption of smoking bans in other units. But our data could be used to examine several other aspects of policy frames, such as whether their mix (e.g., the ratio of different frames or another composite measure) varies over time, whether the topics used focus less on economic consequences over time, and whether states exhibit the same topics that are found in similar states.

# 3 Methodology

## 3.1 Data Sources and Preprocessing

Our analysis of policy frames as a part of the diffusion process concentrates, as noted earlier, on the adoption of antismoking policies in US states. US states historically have had considerable autonomy in public health areas, and smoking restrictions are no exception. Although smoking-related issues are often discussed at the national level (McCann, Shipan and Volden, 2015), few laws have been passed at this level in the US; rather, the vast majority of policymaking has taken place within the states. Thus, the topic of anti-smoking laws provides an excellent forum for examining the process of diffusion.

The time period we examine in the US begins in 1996, which is two years before the first statewide smoking ban was adopted in California.[4] To analyze public discussions and identify policy frames within a state, we rely on articles published in the newspapers listed in Appendix A1. Currently we have processed articles from 50 newspapers in the US covering 47 states, but the full construction of the newspaper corpus is still being completed. The final corpus eventually will include one of the largest newspapers in terms of circulation for every state. We use print media rather than television or radio programs partly for technical reasons but especially because they generally report more extensively on political matters than do on-air media (Druckman, 2005, 469).

We retrieved newspaper texts using a simple, broad keyword search[5] from different database

[4]Debates on smoking bans go back at least to the introduction of the first smoke-free spaces in the 1980s. The Minnesota Clean Indoor Air Act, for example, called for a partial smoking ban in bars and restaurants as early as 1975. However, the analysis requires significant public debates associated with highly visible events.

[5]The keyword string for the different newspaper databases was an adaptation of "tobacco OR non-smoking OR anti-smoking OR smoking OR cigar! OR (lung AND cancer) OR smoker" in the three languages French, German and English.

providers. Then we split the texts into paragraphs of a similar length,[6] which produced a corpus containing 4,134,329 paragraphs for the US. A manual evaluation of a random sample of paragraphs revealed a very low share of paragraphs actually covering smoking bans. This is due to the looseness of our keyword search, aimed at minimizing the number of articles of smoking bans escaping our search. Therefore, we relied on the crowd-sourcing platform Crowdflower[7] to annotate a sample of 10,000 paragraphs as relevant or irrelevant.[8] Following the crow-sourcing guidelines for political science content analyses by Benoit et al. (2016), we found that the crowd annotation produces comparable results with three expert codings. Relevant paragraphs are those containing information on smoking restrictions—that is, bans or limits on smoking in public places or specific workplaces. This definition includes statements about any kind of restriction of smoking ("smoking ban") in public places or businesses introduced through legislative action, executive action, or other democratic actions (e.g., direct-democratic processes). By contrast, we coded as irrelevant paragraphs discussing, for example, smoking bans introduced by private actors (e.g., companies, businesses), or bans of specific tobacco products (e.g., mentholated cigarettes).

Using the information gained by manual coding, we then classified all paragraphs in our corpus as relevant or irrelevant using a text classifier trained with the Python module `scikit-learn`. Prior to the estimation we pre-processed all documents with standard procedures such as text segmentation into paragraphs and sentences, tokenizing, removal of punctuation, collapsing of $n$-word geographical names such as "New York" to one token ("New_York"), as well as lemmatizing, part-of-speech tagging and converting all words to lowercase (Hopkins and King, 2010). For the US newspapers, a kernel ridge regression has proven to be the most effective classifier for our task (see Table 1 for recall[9] and precision[10] on the held-out set of about 10% of the training data). It outperformed a support vector machine, a multinomial naïve bayes classifier as well as any ensemble of all three classifiers.

Moreover, most classification runs we tested agreed with an overall F1-Score of 0.80 or higher—a further sign for the consistency and thus reliability of the classification (Collingwood and Wilkerson, 2012). Therefore, we are confident that our estimations reveal the general trend in the newspapers' coverage of smoking bans. In the end, this filter produced a corpus of 53,526 paragraphs.

The specific form of the keyword string depends on the options available for Boolean operators and truncation wildcards.

[6] The original paragraph structure of the documents was kept, but paragraphs with fewer than 150 tokens were collapsed until the collapsed paragraph exceeded 150 tokens. This ensures the basic comparability of the texts from different newspapers.

[7] `https://www.crowdflower.com/`.

[8] We thank Slava Mikhaylov for suggesting this strategy.

[9] Recall is the fraction of relevant documents that are retrieved.

[10] Precision is the fraction of retrieved documents that are relevant.

Table 1: *Quality of the classification filters.*

| | Precision | Recall | N held-out set |
|---|---|---|---|
| irrelevant | 0.98 | 1.00 | 1838 |
| relevant | 0.98 | 0.70 | 142 |
| average | 0.98 | 0.95 | 1980 |

### 3.2 Estimation

We identify policy frames inductively with a structural topic model (STM) (Roberts et al., 2014; Roberts, Stewart and Airoldi, 2016), which produces estimates document-topic and word-topic probabilities (Roberts, Stewart and Airoldi, 2016; Roberts, Stewart and Tingley, 2014). It builds on well-established generative topic models, such as the Latent Dirichlet Allocation (LDA) (Blei, Ng and Jordan, 2003). The LDA is a mixed-membership model, meaning that it assumes that each document consists of a mixture of topics (Grimmer and Stewart, 2013, 283–285). This assumption is consistent with the strategy used by Baumgartner, De Boef and Boydstun (2008) to code the component parts of frames manually. Concretely, the LDA is a hierarchical model in which a document's $i$ proportion of topics has a common prior drawn from a Dirichlet distribution:

$$\pi_i \sim \text{Dirichlet}(\alpha).$$

Then, the topic of the $j$-th word in the $i$-th document is drawn from a multinomial distribution:

$$\tau_{ij} \sim \text{Multinomial}(\pi_i).$$

Finally, words are drawn from a multinomial distribution where $\theta_{\tau_{ij}}$ is the probability of drawing the $j$-th word for the $j$-th document, conditional on topic $\tau$:

$$w_{ij} \sim \text{Multinomial}(\theta_{\tau_{ij}}).$$

The STM's major innovation is that the prior distribution of topics can be influenced by covariates (Roberts et al., 2014; Roberts, Stewart and Airoldi, 2016):

$$\pi_i \sim \text{LogisticNormal}(X\beta, \Sigma).$$

Furthermore, covariates can also be specified for the word distribution over topics, that is, not only the probability of topics within documents, but also that of words within topics. For instance, this would allow us to see how the language used in a given topic changes as a function of covariates. We will consider this useful option in future work.

Our analysis includes a range of covariates: (1) monthly trend variables with a B-spline of order 10, (2) newspaper IDs and their ideological "slant" (Gentzkow and Shapiro, 2010), (3) the "sentiment" of a paragraph, (4) the percentage of smokers in the state where the newspaper is based and whether tobacco is produced in the state, (5) whether Democrats or Republicans form a unified government in a state, (6) the enactment of smoking bans in a state, (7) the number of months before and after the enactment of smoking bans, and (8) the share of neighboring states having enacted smoking bans ("spatial lag"). For variables 6–8, we consider smoking bans in seven areas: restaurants, bars, government worksites, private worksites, hotels, malls, indoor arenas. The "spatial lag" is computed based on the share of these seven areas in which a state has enacted smoking bans.

For the sentiment analysis, we use an adapted implementation of `word2vec` (Mikolov and Dean, 2013), which learns and aggregates term similarities through a shallow neural network process. The implementation we use is adapted to documents (`doc2vec`), which allows us to consider paragraph and sentence structures as well (Qiu, 2015). We build a `doc2vec` model using our smoking ban paragraphs and 50,000 IMDB movie reviews labelled as positive or negative (Maas et al., 2011). Tested on another 12,500 movie reviews, we achieve an accuracy of 84% for a binary classification into positive and negative. This analysis is in a very early stage. In further iterations, we will obviously have to conduct evaluations on the smoking ban paragraphs directly and, more generally, improve our strategy to measure sentiment.

The spatial lag is the most interesting variable, both substantively and theoretically, and we use it to estimate diffusion effects. In this context, a spatial lag is simply a weighted average of the policies of other states. To construct a spatial lag, we need two pieces of information. First, we need to know when various types of smoking bans were enacted in each the states. We purchased these data from MayaTech's Center for Health Policy and Legislative Analysis, which has already proven to be a highly reliable data source (Shipan and Volden, 2006). Second, we need a connectivity matrix containing information on the relationship between states, specifically, which which states are likely to influence the policies of which other states. While we plan to include more sophisticated indicators in the near future (see Desmarais, Harden and Boehmke, 2015), we simply rely on geographic proximity at the meoment, a catch-all indicator that tends to perform well in practice despite its theoretical bluntness.

When estimating topic models, important decisions pertaining to their the parameters have to be made. In the forefront of the analysis, we therefore conducted a systematic evaluation od how different parameter sets influence the quality of a topic model. In order to objectively measure this quality, we follow O'Callaghan et al. (2015) who use a `word2vec` model to assess the semantic coherence of the most probable word vectors for each topic. In contrast to O'Callaghan et al. (2015), however, we do not only consider the coherence – the similarity of all word pairs in the same topic –, but also the discrimination – the inverse similarity of all word pairs across topics – of a topic model. We ran parameter tests for a candidate range of 5, 10, 15, 20, 25 and 30 topics, six diferent values for $\alpha$[11] as well as five different values for $\eta$[12]. The evaluation yielded an optimal parameter set of 30 topics with $\alpha$=70 and $\eta$=0.01 for the US newspapers (see Figure A2).

# 4 Results

## 4.1 Topics

We present here the results of a model assuming 30 topics, as explained in Section 3.2. Figure 4.1 show the top-50 words associated with the twelve topics most relevant for our purposes, along with labels that we determined based on those words. The other eighteen topics are shown in Appendix A3. The interpretation of all topics is relatively straightforward and their connection with smoking bans quite clear. We conclude that our model identifies relevant and meaningful topics, surprisingly so considering that they were produced purely inductively, without human input.

## 4.2 Validation

To validate the output of the structural topic models, we consider a few correlations that, while theoretically not particularly interesting, help us to assess the plausibility of the results. First, Figure 4.2 shows the passage of legislation was discussed much more frequently during the month in which smoking bans were enacted than in other months, which of course is what we would expect. Second, Figure 4.2 shows that the percent of smokers within a state strongly correlates with four topics: health, individual

[11] 1, 10, 20, 50, 70 and 100 divided by the number of topic in the repsective run; $\alpha$ is the prior for the topic-document distribution.

[12] 1, 0.1, 0.01, 0.001 and 0.0001; $\eta$ is the prior for the topic-word distribution.

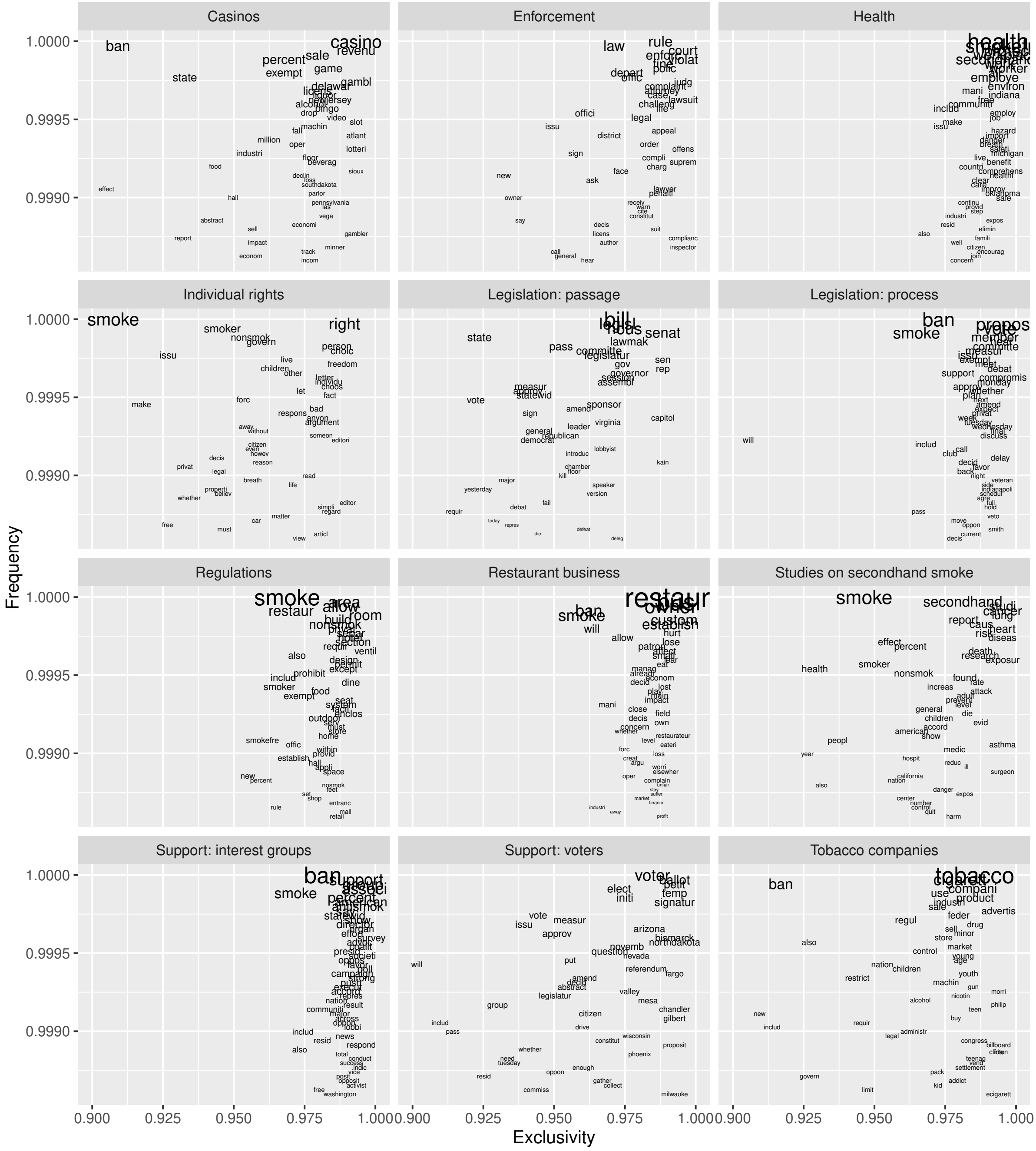


Figure 1: *Top-50 words for the twelve most relevant topics.*

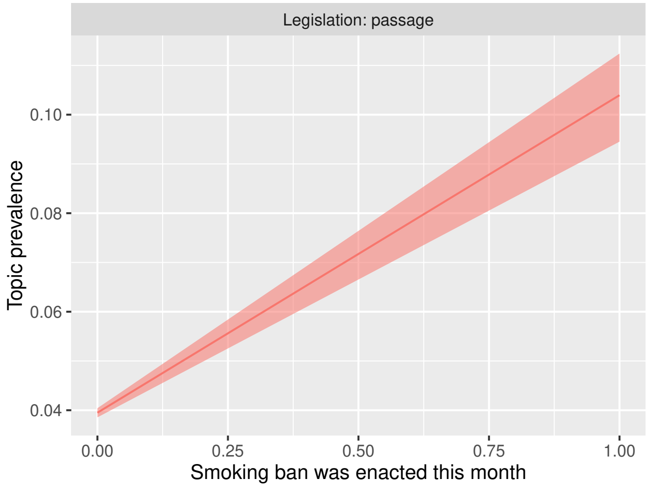


Figure 2: *Validation of the structural topic model: The passage of legislation is discussed more often during months in which legislation was passed.*

rights, regulations, and restaurant business. All these topics are more important when a larger share of the population smokes. The health benefits of smoking restrictions are larger where more people benefit from them; concerns about restrictions of individual freedom are more widespread if more people see their freedom to smoke restrained; detailed provisions such as ventilation or the establishment of smoking areas are more salient when a larger share of the population will potentially continue to smoke after the passage of smoking restrictions; and questions surrounding the (potentially negative) consequences of smoking bans for the restaurant business loom larger in states where more patrons are smokers. We conclude from these results that the structural topic models identify correlations that make intuitive sense. This allows us to proceed with more confidence to the interpretation of the main findings in the next section.

### 4.3 The Diffusion of Policy Frames

The argument of this paper is that the frames used to discuss smoking bans are related to the presence of the policy in neighboring states. Figure 4.3 provides direct evidence of this phenomenon by showing how the prevalence of the twelve main topics varies as a function of the share of neighboring states that have enacted smoking bans.

Not all topics are correlated with the policies of other states. *Enforcement*, for instance, is discussed with about the same frequency regardless of how many neighboring states have enacted smoking bans. Enforcement issues are salient in public debates on smoking bans—their frequency is above the baseline

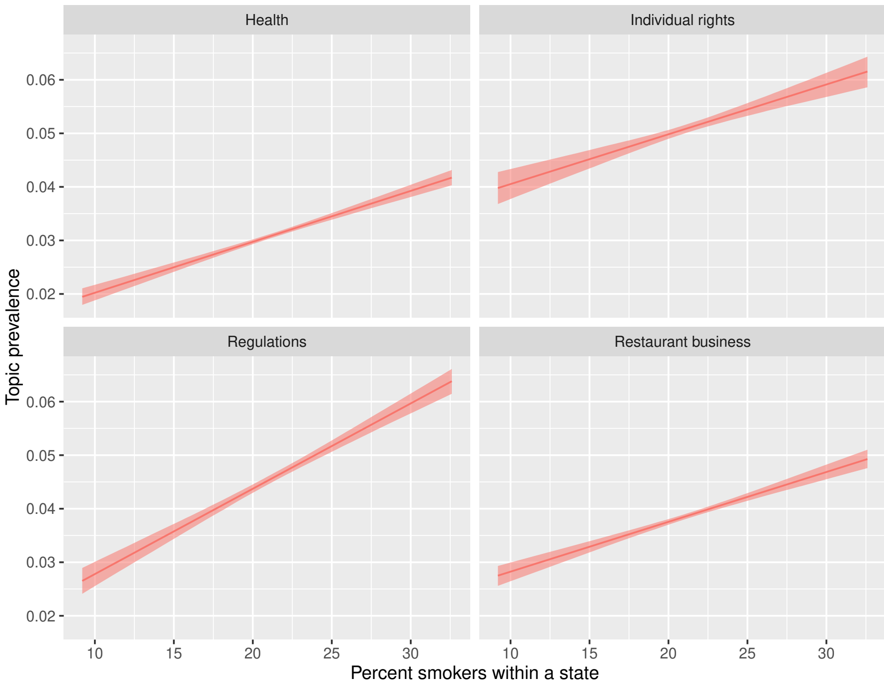


Figure 3: *Validation of the structural topic model: The health implications of smoking bans, their compatibility with individual rights, specific regulations such as separate rooms for smokers, and the consequences of smoking bans for the restaurant business are discussed more frequently in states with many smokers than in states with few.*

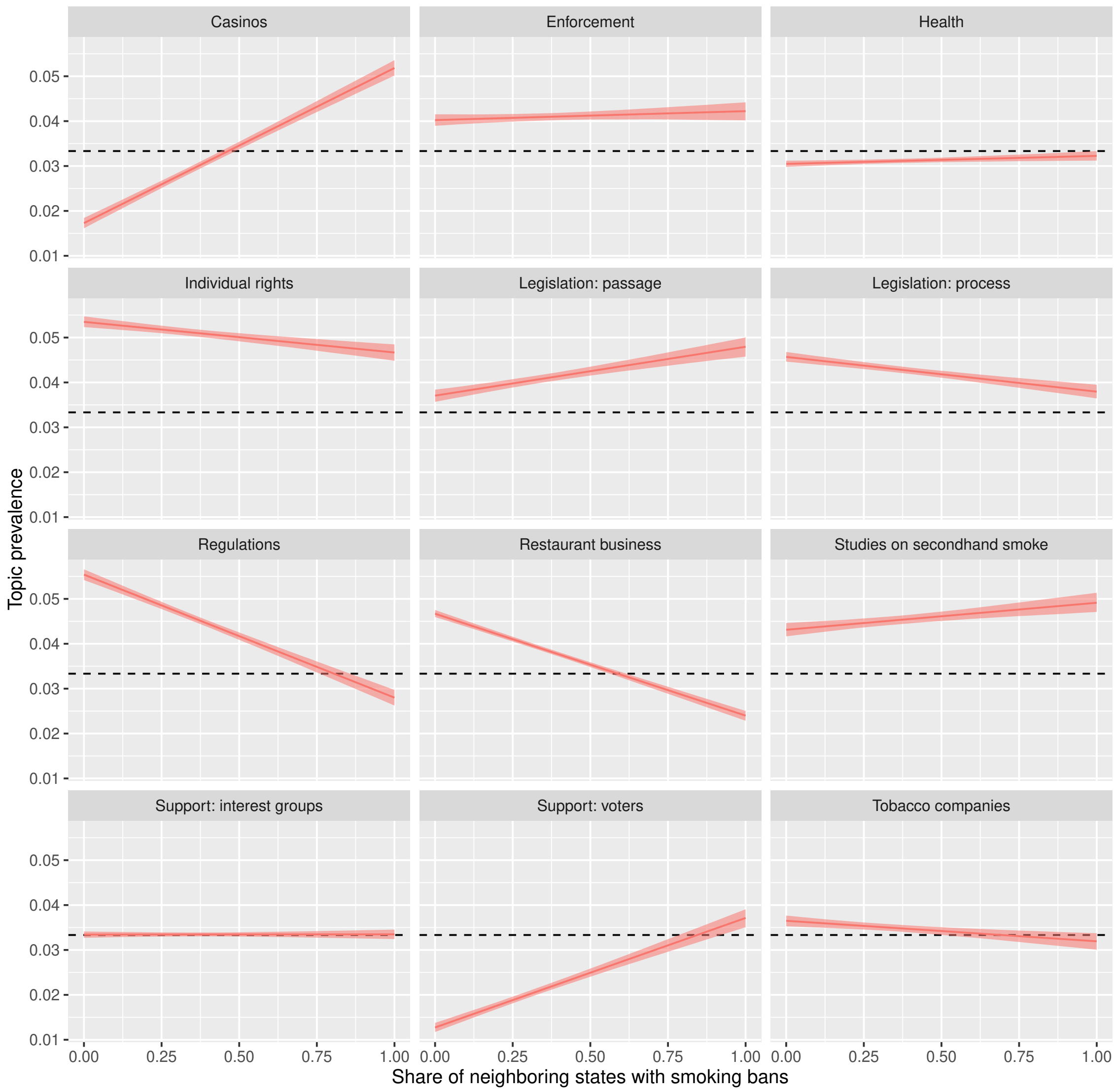


Figure 4: *Some topics are more prevalent when smoking bans are widespread among neighboring states, while other topics are more prevalent when few neighboring states have enacted smoking bans. Not all topics are correlated with the presence of smoking bans in neighboring states. The horizontal line shows baseline topic prevalence.*

value of topic prevalence—but their salience does not increase or decrease when more information about the ease of enforcing smoking bans becomes available in neighboring states. The same holds for other topics, such as *Health* or *Support: interest groups*.

By contrast, the prevalence of some topics changes strongly depending on the policies of neighboring states. The correlation is strong—and negative—for *Restaurant business*, which is a topic that tends to be hotly debated whenever smoking restrictions first appear on the political agenda. A common worry is that prohibiting smoking in restaurants might harm their business. Although we have not conducted a satisfactory sentiment analysis yet, it is clear from the words in Figure 4.1 (e.g. "lose", "hurt", "fear", "affect") that the tone of the texts for this topic tends to be negative. Thus, Figure 4.3 shows that concerns about a negative impact of smoking bans for the restaurant business decrease significantly when the experience of more neighboring states becomes available, arguably showing that the policy is not harmful to restaurants. The exact opposite happened with *Casinos*. Similar to *Restaurant business*, the words in Figure 4.1 suggest a negative tone ("drop", "fall", "decline"). In contrast to *Restaurant business*, however, *Casinos* becomes more salient when more neighboring states enact smoking bans, suggesting that their experience points to negative consequences for the casino business. These findings connect directly with the idea of policy learning, namely, that beliefs about the consequence of a policy (i.e. whether it harms business or not) are updated based on what can be seen elsewhere. Figure 4.3 supports this notion.

But learning is not just about policy outcomes, it is also about political outcomes (Gilardi, 2010). The topic *Support: voters* identifies voters' involvement in the decision-making process. Although the tonality of the topic is difficult to infer from the top words and is possibly ambivalent, Figure 4.3 shows that the topic's prevalence increases sharply when smoking bans become more widespread among neighboring states. Voters' views hardly receive any attention when no other state has enacted smoking bans, but they become a much more prominent topic when neighboring states start to pass smoking restrictions. By contrast, another dimension of political support, that of interest groups (*Support: interest groups*), seems completely unrelated to other states' policies.

A strong correlation is also apparent for *Regulations*. This topic identifies the technical aspects of smoking bans, such as rules or permits for separate smoking areas, ventilation, exemptions, and so on. Getting these regulations right is important for the implementation of smoking bans and uncertainty surrounding them may worry business owners. Figure 4.3 shows that these issues are very salient when no other state has enacted smoking bans, and much less so when many have. Like for *Restaurant business* or *Casinos*, this finding suggests that the experiences of other states are used to update prior beliefs—in this case, what kind of regulations work best or how difficult it is to bet them right.

Other correlations are less strong but worth discussing. *Legislation: passage* is positively correlated with the presence of smoking bans in other states, while *Legislation: process* is negatively correlated. The former topic identifies the act of passing a bill, while the latter picks up aspects of the decision-making process, including debates, discussions, proposals, the degree of support, amendments, and so on. The positive correlation with *Legislation: passage* likely stems from the diffusion of smoking ban legislation itself—if states are more likely to adopt smoking restrictions when more neighbors do so, this will be reflected in news coverage. The negative correlation with *Legislation: process* suggests that the decision-making process may become less controversial, or less news-worthy, when the policy becomes more widespread.

Finally, the prevalence of *Individual rights* decreases with the share of neighboring states with smoking restrictions. This topic refers to debates on the appropriateness of limiting the individual choice to smoke. It loses significance as the policy spreads, suggesting that it increasingly becomes perceived as adequate from the perspective of personal freedom. This finding points to the more normative components of policy diffusion.

## 5 Conclusion

Policy diffusion is a multi-stage process, but most research has been limited to an examination of only two of these stages—the initial adoption (or adoptions) in some set of states, and then whether future adoptions are influenced by these earlier adoptions. We argue that an intermediary stage is of crucial importance, both because it is affected by earlier adoptions and because it can affect later adoptions. More specifically, it is during this intermediary stage—the second stage of the diffusion process—the emergence of specific policy frames. These frames can plausibly influence the likelihood of adoption, but our interest in this paper is on examining the frames themselves. What frames exist? Do these frames vary over time? And most importantly, are these frames a function of earlier adoptions elsewhere? To the extent that these frames are a function of earlier adoptions, we should recognize them as a critical part of the overall diffusion process.

Our analysis provides a first step toward better understanding how policy frames can diffuse—or more accurately, how the first stage of the diffusion process, in which other states adopt policies, can influence the next stage, in which the relative strength of different policy frames changes. We have put forward a preliminary analysis of the diffusion of the framing of smoking bans in US states based on a structural topic model of over 50,000 paragraphs in 50 American newspapers covering 47 states between 1996 and 2014. Results show that there is variation in the incidence of these frames, as well as connections between these frames and the prevalence of prior adoptions in neighboring states. In particular, as more neighboring states enact smoking bans,

- concerns about their implications for the restaurant business decrease;
- concerns about their implications for casinos increase;
- discussions surrounding detailed regulations such as ventilation requirements or separate rooms for smokers become less prevalent;
- voters' support and involvement in the decision-making process surrounding smoking bans is discussed more frequently;
- the compatibility of smoking restrictions with individual rights loses salience as a topic;
- the passage of legislation is discussed more frequently, while the process by which decisions are made loses salience.

More work remains to be done. Obtaining estimates of whether the newspaper coverage was positive or negative will allow us to ascertain whether not only the frame, but the nature of the frame, varies in response to earlier adoptions. The sentiment analysis we conducted (but not reported in detail) is a first step in this direction. For improving the analysis of sentiment, we are currently applying the crowd-sourced approach put forward by Benoit et al. (2016), which we already have used to code relevant paragraphs for the machine-learning classifier. Secondly, we will improve the construction of the spatial lag by relying on the measures put forward by Desmarais, Harden and Boehmke (2015), which, however, need to be extrapolated for a few years. For now, however, our preliminary analysis has established a foothold for the usefulness of structural topic models and support for the idea that policy frames are an important part of the diffusion process.

## References


Baumgartner, Frank R., Suzanna L. De Boef and Amber E. Boydstun. 2008. *The Decline of the Death Penalty and the Discovery of Innocence*. New York: Cambridge University Press.

Benoit, Kenneth, Drew Conway, Benjamin E Lauderdale, Michael Laver and Slava Mikhaylov. 2016. "Crowd-Sourced Text Analysis: Reproducible and Agile Production of Political Data." *American Political Science Review* forthcoming.

Blei, D. M., A. Y. Ng and M. I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning* 3:993–1022.

Braun, Dietmar and Fabrizio Gilardi. 2006. "Taking 'Galton's Problem' Seriously: Towards a Theory of Policy Diffusion." *Journal of Theoretical Politics* 18(3):298–322.

Collingwood, Loren and John Wilkerson. 2012. "Tradeoffs in Accuracy and Efficiency in Supervised Learning Methods." *Journal of Information Technology & Politics* 9(3):298–318.

Desmarais, Bruce A, Jeffrey J Harden and Frederick J Boehmke. 2015. "Persistent Policy Pathways: Inferring Diffusion Networks in the American States." *American Political Science Review* 109(2):392–406.

Dobbin, Frank, Beth Simmons and Geoffrey Garrett. 2007. "The Global Diffusion of Public Policies: Social Construction, Coercion, Competition, or Learning?" *Annual Review of Sociology* 33:449–472.

Druckman, J. N. 2005. "Media Matter: How Newspapers and Television News Cover Campaigns and Influence Voters." *Political Communication* 22(4):463–481.

Entman, R. M. 1993. "Framing: Toward Clarification of a Fractured Paradigm." *Journal of Communication* 43(4):51–58.

Gentzkow, Matthew and Jesse M Shapiro. 2010. "What Drives Media Slant? Evidence from US Daily Newspapers." *Econometrica* 78(1):35–71.

Gilardi, Fabrizio. 2010. "Who Learns from What in Policy Diffusion Processes?" *American Journal of Political Science* 54(3):650–666.

Gilardi, Fabrizio. 2012. Transnational Diffusion: Norms, Ideas, and Policies. In *Handbook of International Relations*, ed. Walter Carlsnaes, Thomas Risse and Beth Simmons. Second ed. SAGE Publications pp. 453–477.

Graham, Erin R., Charles R. Shipan and Craig Volden. 2013. "The Diffusion of Policy Diffusion Research in Political Science." *British Journal of Political Science* 43(3):673–701.

Grimmer, Justin and Brandon M Stewart. 2013. "Text as data: The promise and pitfalls of automatic content analysis methods for political texts." *Political Analysis* 21:267–297.

Hopkins, Daniel J. and Gary King. 2010. "A Method of Automated Nonparametric Content Analysis for Social Science." *American Journal of Political Science* 54(1):229–247.

Maas, A. L., R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng and C. Potts. 2011. *Learning Word Vectors for Sentiment Analysis*. The 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011).

Maggetti, Martino and Fabrizio Gilardi. 2016. "Problems (and Solutions) in the Measurement of Policy Diffusion Mechanisms." *Journal of Public Policy* 36(1):87–107.

McCann, Pamela J Clouser, Charles R Shipan and Craig Volden. 2015. "Top-Down Federalism: State Policy Responses to National Government Discussions." *Publius* 45(4):495–525.

Mikolov, T., K. Chen G. Corrado and J. Dean. 2013. *Efficient Estimation of Word Representations in Vector Space*. CoRR, abs/1301.3781.

O'Callaghan, D., D. Greene, J Carthy and P. Cunningham. 2015. "An analysis of the coherence of descriptors in topic modeling." *Expert Systems with Applications* 42:5645–5657.

Pacheco, Julianna. 2012. "The Social Contagion Model: Exploring the Role of Public Opinion on the Diffusion of Antismoking Legislation across the American States." *Journal of Politics* 74(1):187–202.

Qiu, L. 2015. *Sentiment Analysis using Doc2Vec*. https://github.com/linanqiu/word2vec-sentiments.

Roberts, M. E., B. M. Stewart and D. Tingley. 2014. *stm: R Package for Structural Topic Models*.

Roberts, M. E., B. M. Stewart, D. Tingley, C. Lucas, J. Leder-Luis, S. Gadarian, B. Albertson and D. Rand. 2014. "Structural topic models for open-ended survey responses." *American Journal of Political Science* 58:1064–1082.

Roberts, Margaret E., Brandon M. Stewart and Edoardo M. Airoldi. 2016. "A model of text for experimentation in the social sciences." *Journal of the American Statistical Association* forthcoming.

Shipan, Charles R. and Craig Volden. 2006. "Bottom-Up Federalism: The Diffusion of Antismoking Policies from U.S. Cities to States." *American Journal of Political Science* 50(4):825–843.

Shipan, Charles R. and Craig Volden. 2008. "The Mechanisms of Policy Diffusion." *American Journal of Political Science* 52(4):840–857.

Shipan, Charles R. and Craig Volden. 2014. "When the Smoke Clears: Expertise, Learning and Policy Diffusion." *Journal of Public Policy* 34(3):357–387.

Simmons, Beth, Frank Dobbin and Geoffrey Garrett. 2006. "Introduction: The International Diffusion of Liberalism." *International Organization* 60(4):781–810.

# A1 Newspaper corpus

| *Newspaper* | *State* | *Articles* | *Paragraphs* | *Filtered* |
|---|---|---|---|---|
| Albuquerque Journal | NM | 4,953 | 25,464 | 750 |
| Argus Leader | SD | 3,801 | 25,339 | 887 |
| Arizona Republic | AZ | 9,405 | 44,455 | 1,590 |
| Atlanta Journal-Constitution | GA | 23,281 | 114,843 | 1,196 |
| Austin American-Statesman | TX | 12,573 | 86,686 | 1,251 |
| Birmingham News* | AL | 1,914 | 9,000 | 105 |
| Bismarck Tribune | ND | 10,251 | 40,867 | 1,971 |
| Boston Globe | MA | 19,337 | 112,465 | 2,128 |
| Charleston Gazette-Mail | WV | 18,228 | 116,099 | 1,428 |
| Chicago Tribune | IL | 31,855 | 157,102 | 3,225 |
| Courier-Journal | KY | 10,593 | 71,887 | 2,622 |
| Daily News | NY | 14,202 | 60,828 | 571 |
| Oklahoman | OK | 12,250 | 44,793 | 741 |
| Denver Post | CO | 13,088 | 79,843 | 994 |
| Deseret News | UT | 15,884 | 58,817 | 702 |
| Des Moines Register | IA | 5,750 | 41,160 | 714 |
| Detroit Free Press | MI | 11,309 | 115,380 | 694 |
| Hartford Courant | CT | 14,821 | 83,980 | 517 |
| Honolulu Star-Advertiser | HI | 1,465 | 8,282 | 147 |
| Idaho Falls Post Register | ID | 2,082 | 11,083 | 125 |
| Indianapolis Star | IN | 11,432 | 92,001 | 2,707 |
| Las Vegas Review-Journal | NV | 9,430 | 56,605 | 779 |
| Los Angeles Times | CA | 29,597 | 196,061 | 1,281 |
| Journal Sentinel | WI | 16,040 | 81,146 | 805 |
| New York Times | NY | 53,411 | 344,898 | 1,684 |
| Omaha World-Herald | NE | 12,295 | 72,506 | 1,708 |
| Philadelphia Inquirer | PA | 18,975 | 105,861 | 1,080 |
| Portland Press Herald | ME | 5,374 | 27,796 | 540 |
| Providence Journal | RI | 15,264 | 89,549 | 967 |
| North Jerey Record | NJ | 19,453 | 95,395 | 1,191 |
| Richmond Times-Dispatch | VA | 23,237 | 141,295 | 1,323 |
| Star Tribune | MN | 13,693 | 120,220 | 1,494 |
| St.Louis Post-Dispatch | MI | 27,516 | 137,830 | 2,203 |
| Tampa Bay Times | FL | 22,369 | 162,254 | 1,592 |
| Baltimore Sun | MD | 14,264 | 86,647 | 1,957 |
| Billings Gazette | MT | 235 | 1,416 | 101 |
| Burlington Free Press | VT | 1,938 | 10,607 | 365 |
| Clarion-Ledger | MS | 3,206 | 17,005 | 370 |
| News Journal | DE | 5,426 | 31,177 | 1,218 |
| Oregonian* | OR | 3,406 | 17,154 | 430 |
| Plain Dealer* | OH | 3,394 | 19,672 | 268 |
| Seattle Times | WA | 16,820 | 79,862 | 647 |
| Tennessean | TN | 5,475 | 36,611 | 470 |
| Times-Picayune* | LA | 3,600 | 17,776 | 448 |
| Union Leader* | NH | 975 | 3,944 | 57 |
| Topeka Capital-Journal | KS | 5,976 | 32,294 | 857 |
| USA Today | NY | 11,246 | 59,637 | 460 |
| Wall Street Journal | NY | 22,971 | 139,448 | 658 |
| Washington Post | DC | 58,495 | 501,552 | 2,098 |
| Wilmington Star-News | NC | 6,863 | 34,211 | 720 |
| Wyoming Tribune Eagle | WY | 2,024 | 13,526 | 690 |
| Total | | 681,442 | 4,134,329 | 53,526 |

* For these newspapers, several years of coverage could not be retrieved.

Table A1: *Newspaper corpus.*

## A2 Topic model coherence

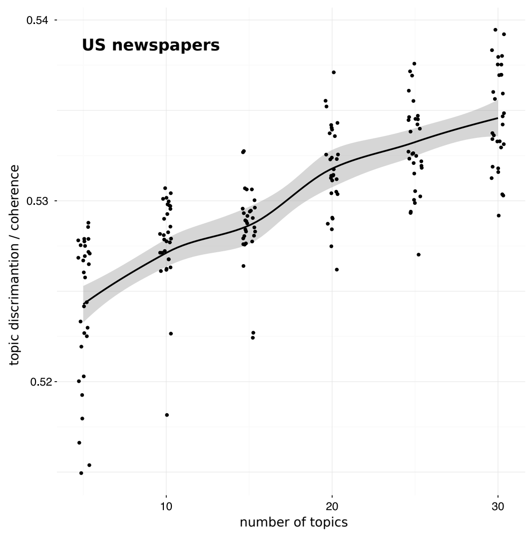


Table A2: *Word2vec topic coherence and discrimination averages for varying numbers of topics, alpha priors and eta priors.*

# A3 Other topics